

# Genetic Variation and Association Mapping of Protein Concentration in Brown Rice Using a Diverse Rice Germplasm Collection

Rolfe J. Bryant,<sup>1</sup> Aaron K. Jackson,<sup>1</sup> Kathleen M. Yeater,<sup>2</sup> Wengui G. Yan,<sup>1</sup> Anna M. McClung,<sup>1</sup> and Robert G. Fjellstrom<sup>1,3</sup>

## ABSTRACT

Cereal Chem. 90(5):445–452

Protein is the second most abundant constituent in the rice grain next to starch. Association analysis for protein concentration in brown rice was performed using a “mini-core” collection, which represents the germplasm diversity found in the USDA rice world collection. Protein concentration was determined in replicated trials conducted in two southern U.S. locations, and association mapping was performed by using 157 genomewide DNA markers. Protein concentration ranged from 5.4 to 11.9% among the 202 accessions. Protein variation owing to accession and accession × location interaction were highly significant. Ample variation was seen within each subpopulation by ancestry, as well as within the

14 geographic regions where the accessions originated. Accessions from Eastern Europe had the highest level of protein. Ten markers on eight chromosomes were significantly associated with protein concentration. Five of these markers occurred near known protein precursor genes or quantitative trait loci, and the other five markers were novel for the association with protein concentration in rice. The germplasm and genetic markers identified in this study will assist breeders in developing cultivars tailored for applications requiring specific protein concentration in the rice grain. The research results contribute to the potential discovery of novel rice storage protein pathways in the endosperm.

Rice (*Oryza sativa* L.) is an important source of nutrition and energy for the majority of the global population. Next to starch, protein is the second most abundant constituent of the rice grain. However, the protein content of rice is the lowest of all the cereals (Eggum 1969, 1977, 1979; Wolff 1982). Rice protein concentration is affected by climate and agronomic conditions and is non-uniformly distributed in the rice grain, with the greatest concentration in the aleurone layer of the endosperm (Little and Dawson 1960). Besides the nutritional value, rice protein influences the texture, pasting properties, and sensory characteristics of rice (Champagne et al 2007, 2009; Bryant et al 2012). It has been reported that under certain conditions the protein concentration of rice could be increased 1.8-fold by increasing nitrogen fertilization (Champagne et al 2009).

Compared with the starch content, total protein concentration is relatively low and constitutes only 7% of the rice endosperm, with 95% of the protein being contained within protein bodies (Shih 2004). Rice seed storage proteins are divided into four main groups based on their solubility. Glutelin is the most abundant protein in the endosperm and makes up 60–80% of the total protein content. To date, 15 glutelin genes have been identified in rice (Kawakatsu et al 2008). Prolamins are the second most abundant seed storage protein, making up 20–30% of the total protein, and 34 genes encoding prolams have been identified (Kawakatsu et al 2010). Albumin and globulin make up the remaining protein components in the endosperm. Six genes encoding albumins and three genes encoding globulins have been identified (Rice Genome Annotation Project, rice.plantbiology.msu.edu).

Understanding the underlying genetic pathways involved in rice seed storage proteins and manipulating rice protein concentration in the endosperm have been of considerable importance for the biopharmaceutical and nutraceutical industries in recent years. Expression of foreign recombinant proteins in rice endosperm has

been used to increase vitamin A concentration in the well-known Golden Rice (Paine et al 2005), in medical studies to produce recombinant proteins capable of reducing hypertension (Tada et al 2003), and for producing human serum albumin (He et al 2011). In addition, modifying protein concentrations for traditional uses, such as increasing nutritional value or enhancing grain quality, by understanding the naturally occurring mechanisms through which protein concentration is regulated in the endosperm has been proposed to offer a wider range of applications for transgenic uses. The downregulation of endogenous seed storage proteins can allow space for the accumulation of more recombinant proteins (Tada et al 2003). Conversely, the ability to upregulate protein concentration can enable greater production of desired proteins in transgenic lines using native rice seed storage promoters (Qu and Takaiwa 2004).

To better identify and understand the genomic locations controlling protein concentration in the rice grain, we conducted a genomewide association study by using a genetically diverse collection of rice lines. The USDA world collection of rice contains over 18,000 accessions from 115 countries (Bockelman et al 2003). To improve the efficiency of evaluating and characterizing the phenotypic traits and genetic diversity of rice found in the world collection, a representative core subset (1,794 accessions) was developed by using a stratified random sampling from the USDA world collection (Yan et al 2007). Subsequently, from this subset, a “mini-core” collection of 217 accessions was developed for a comprehensive and extensive characterization of the phenotypic and genotypic traits in replicated experiments (Agrama et al 2009). This mini-core collection includes representatives of five rice subpopulations: indica, aus, temperate japonica, tropical japonica, and aromatic cultivars and landraces, plus the related species *O. rufipogon* and *O. glaberrima* (Li et al 2010). The broad international origins of the mini-core accessions from 76 countries contribute to one of the most diversified collections in rice, which is valuable for association mapping to identify genetic marker differences that are correlated with specific traits. Association mapping has been successfully employed to identify numerous and novel gene regions controlling economically important cereal crop traits, thus proving to be a powerful alternative or complement to biparental inheritance studies of traits (Kraakman et al 2004; Zhang et al 2005; Bresegheello and Sorrells 2006; Ravel et al 2006; Agrama et al 2007; Ersoz et al 2007; Famoso et al 2011; Zhao et al 2011).

The objectives of this study were 1) to determine the genetic variability for protein concentration in the mini-core subset of the

<sup>1</sup>U.S. Department of Agriculture—Agricultural Research Service (USDA-ARS), Dale Bumpers National Rice Research Center, 2890 Hwy 130 E., Stuttgart, AR 72160, U.S.A. Mention of a trademark or proprietary product in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

<sup>2</sup>USDA-ARS, Southern Plains Area, College Station, TX 77840, U.S.A.

<sup>3</sup>Corresponding author. Phone: (870) 672-9300, ext. 223. Fax: (870) 673-7581. E-mail: bob.fjellstrom@ars.usda.gov

USDA world rice collection; 2) to determine the relationship of protein content with subpopulation structure and geographic region of accession origin; and 3) to determine the genetic markers associated with protein concentration by using linkage disequilibrium mapping techniques.

## MATERIALS AND METHODS

### Mini-Core Collection

Based on their ability to produce fully mature seeds in the southern U.S. environment, 202 accessions were selected from the mini-core collection (Agrama et al 2009). The selected accessions were grown in field plots near the Dale Bumpers National Rice Research Center, Stuttgart, Arkansas (AR), and the Rice Research Unit, Beaumont, Texas (TX), in 2009. The seeds used for sowing came from a single plant selection characteristic of the germplasm accession from which it originated. A randomized complete block design was used with two replications per location. The samples from each 0.3 × 1.5 m plot were harvested by hand, threshed and cleaned with standard field equipment (Almaco, Nevada, IA, U.S.A.), and dried to approximately 12% moisture with a forced-air drier. Rough rice samples from the plots were stored at 4°C and 50% relative humidity until dehulling and grinding.

### Protein Assay

Brown rice was produced by dehulling rough rice with a Satake testing husker (Satake Engineering Co., Uino Taito-Ku, Tokyo, Japan) and then ground in a Cyclotech grinder (Foss North America, Eden Prairie, MN, U.S.A.) with a 0.5 mm screen. Nitrogen concentration was determined on ground brown rice with a Leco FP-2000 nitrogen analyzer (Leco Corporation, St. Joseph, MI, U.S.A.) and reported as percent protein by converting the nitrogen value to protein concentration with 5.95 as a conversion factor (AACC International Approved Method 46-30.01). All analytical measurements were conducted in duplicate.

### Genotyping

The mini-core accessions were genotyped with 156 simple sequence repeat markers (marker sequence and genomic location information available online at the Gramene website, [www.gramene.org](http://www.gramene.org)), three nucleotide insertion/deletion markers (RID 12, Sweeney et al 2006; Pi-ta, Wang et al 2010; and an insertion/deletion marker found in the first intron of the rice *Waxy* gene), and six single nucleotide polymorphism markers (EX1 G/T, EX6, and EX10 from the rice *Waxy* gene, Chen et al 2010; and three *Alk* gene single nucleotide polymorphisms, Bao 2006). Of these markers, 157 were subsequently used for association mapping. Marker names and positions are shown in Table I.

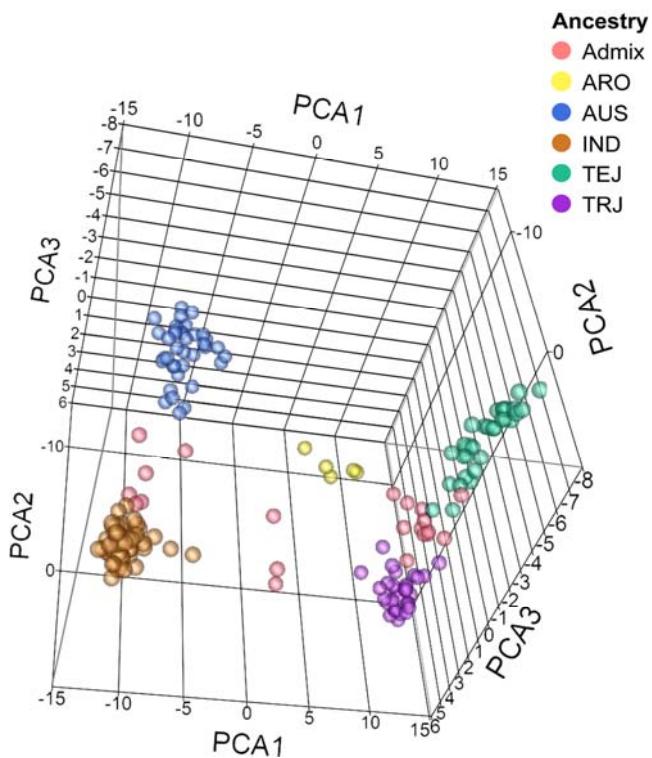
**TABLE I**  
**Name and Location of DNA Markers Used in Association Mapping of Protein Concentration in Mini-Core Rice Accessions<sup>a</sup>**

Marker	Ch	Mb	Marker	Ch	Mb	Marker	Ch	Mb	Marker	Ch	Mb
RM495	1	0.2	RM55	3	29.0	AP5659-1	6	10.4	LJSSR1	9	21.5
RM6324	1	2.4	RM8203	3	31.4	RM541	6	19.5	RM6971	9	21.9
RM283	1	5.3	RM468	3	32.7	RM3827	6	22.3	RM189	9	22.1
RM490	1	6.7	RM514	3	35.3	RM162	6	24.0	RM6816	9	22.2
RM600	1	9.5	RM518	4	2.0	RM5371	6	25.8	RM245	9	22.3
RM3412	1	11.6	RM5953	4	9.4	RM340	6	28.6	RM1026	9	22.6
RM312	1	14.9	RM3317	4	13.6	OSR21	6	30.9	RM1013	9	22.8
RM11008	1	17.9	RM471	4	18.8	RM3394	7	0.7	RM216	10	5.4
RM5	1	24.0	RM1359	4	19.9	RM427	7	2.7	RM25289	10	11.6
RM488	1	24.8	RM3558	4	22.8	RM5711	7	3.1	RM25421	10	14.7
RM237	1	26.8	RM273	4	23.8	RM125	7	5.5	RM271	10	16.6
RM1231	1	29.5	AL606682-1	4	27.7	RID12 <sup>b</sup>	7	6.1	RM171	10	19.1
RM302	1	32.9	RM317	4	29.1	RM1186	7	9.0	RM25766	10	20.1
RM1339	1	38.2	RM3217	4	30.1	RM2	7	16.0	RM484	10	21.1
RM431	1	38.9	RM8217	4	32.7	RM11	7	19.3	RM333	10	22.4
RM154	2	1.1	RM124	4	34.7	RM455	7	22.4	RM7203	11	1.1
RM279	2	2.9	RM507	5	0.1	RM505	7	24.5	RM6544	11	3.9
RM555	2	4.3	RM5796	5	1.3	RM118	7	26.6	RM116	11	5.7
RM145	2	7.7	RM413	5	2.2	RM1335	7	28.3	RM26333	11	7.6
RM452	2	9.6	RM3777	5	4.2	RM408	8	0.1	RM536	11	9.0
RM424	2	11.7	RM169	5	7.5	RM152	8	0.7	RM5857	11	11.9
RM561	2	18.8	RM509	5	16.3	RM6863	8	2.0	RM287	11	16.7
RM341	2	19.4	RM146	5	18.0	RM1148	8	3.7	RM5349	11	19.2
RM475	2	20.4	RM161	5	20.8	RM25	8	4.4	RM254	11	23.7
RM5427	2	21.6	RM5401	5	22.2	RM22559	8	5.7	RM1233	11	26.5
RM450	2	28.6	RM178	5	25.1	RM44	8	11.8	RM224	11	27.2
RM530	2	30.5	RM334	5	28.5	RM404	8	15.4	RM19	12	2.4
RM250	2	32.8	RM133	6	0.2	RM23000	8	17.8	RM247	12	3.2
RM208	2	35.1	Ex1G/T <sup>c</sup>	6	1.764	RM284	8	21.0	RM5746	12	5.1
RM231	3	2.5	Int1InDel <sup>b</sup>	6	1.764	RM149	8	24.7	RM7003	12	6.8
RM489	3	4.3	Ex6 <sup>c</sup>	6	1.767	RM433	8	25.8	Pi-ta <sup>b</sup>	12	10.6
RM14643	3	7.1	Ex10 <sup>c</sup>	6	1.768	RM447	8	26.5	RM7102	12	13.2
RM232	3	9.8	RM190	6	1.8	RM316	9	1.1	RM277	12	18.3
RM3562	3	11.7	RM510	6	2.8	RM23869	9	6.3	RM463	12	22.1
RM338	3	13.2	AP5625-1	6	5.3	RM24011	9	9.4	RM3726	12	23.2
RM15123	3	15.8	AP5625-2	6	5.3	RM105	9	12.5	RM3739	12	25.0
RM3400	3	17.2	TQNG <sup>c</sup>	6	6.7	RM409	9	14.4	RM1300	12	26.0
RM5864	3	22.4	NPBRC <sup>c</sup>	6	6.8	RM434	9	15.7			
RM5626	3	24.9	ALK <sup>c</sup>	6	6.8	RM553	9	19.3			
RM6736	3	27.3	RM527	6	9.9	RM215	9	21.2			

<sup>a</sup> All markers are simple sequence repeat based, unless otherwise noted. Ch = rice chromosome number, and Mb = position of marker on rice chromosome referenced to 2006 annotated Nipponbare sequence.

<sup>b</sup> Nucleotide insertion/deletion marker.

<sup>c</sup> Single-nucleotide polymorphism marker.



**Fig. 1.** Three-dimensional diagram of mini-core accessions plotted on the first three principal component coordinates explaining 90.7% of the marker similarity variation between accessions. ARO = aromatic; IND = indica; TEJ = temperate japonica; TRJ = tropical japonica; and admix = admixture (genetically interbred).

**TABLE II**  
Variance Component Estimates of Random Effect Factors  
for Protein Content Among Accession Entries in Mini-Core  
Grown in Two U.S. Locations (Arkansas and Texas)

Parameter	Estimate <sup>a</sup>	SE	z Value <sup>b</sup>	Total (%) <sup>c</sup>
Accession (A)	0.8233	0.1248	6.60*	40.05
Location (L)	0.5181	0.7346	0.71	25.21
A × L	0.4592	0.0639	7.18*	22.34
Replication	0.0001	0.0011	0.01	0.00
Residual	0.2549	0.0181	14.07*	12.40
Total	2.0555	...	...	100.00

<sup>a</sup> Variance component estimates are representative of proportions of the total variance and are summed to calculate the total variance.

<sup>b</sup> Wald tests of variance components; \* indicates  $P < 0.01$ .

<sup>c</sup> Each parameter's proportion of the total variance.

DNA from the leaf tissue of each mini-core accession was extracted following two methods, a CTAB method as described by Hulbert and Bennetzen (1991) and a rapid DNA extraction method described by Xin et al (2003). Polymerase chain reaction (PCR) marker amplifications were performed as described by Bryant et al (2011) in MJ Research thermal cyclers (Bio-Rad, Hercules, CA, U.S.A.) with PCR primers that were labeled with a fluorescent dye (6FAM, NED, or Hex; Applied Biosystems, Foster City, CA, U.S.A., or Integrated DNA Technologies, Coralville, IA, U.S.A.). The marker amplification reactions were scored for their product sizes (alleles) with an ABI 3700 capillary-electrophoresis genetic analyzer according to the manufacturer's specifications (Applied Biosystems).

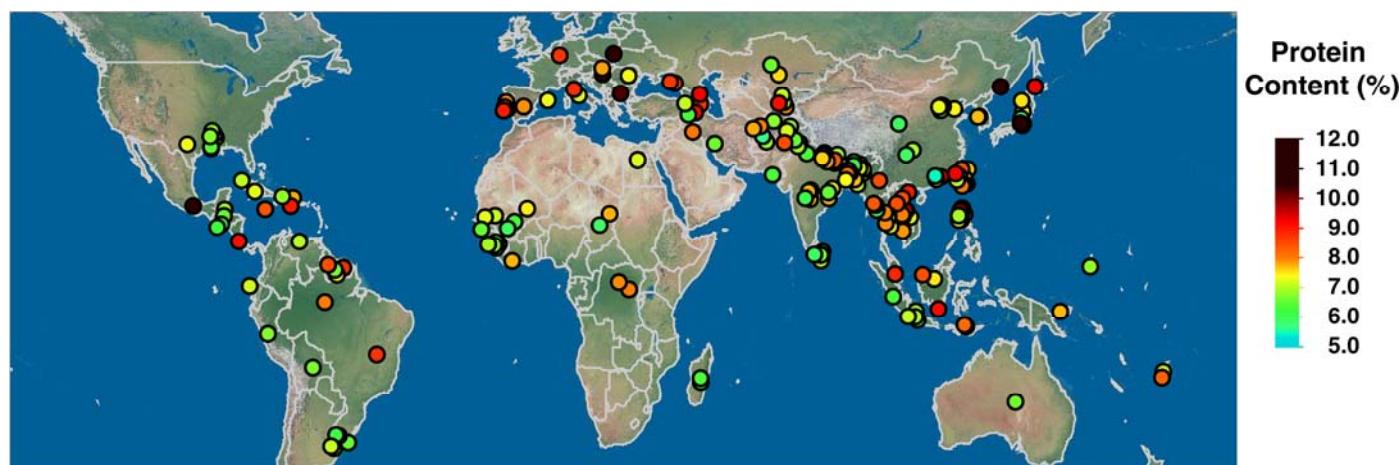
#### Statistical and Population Structure Analyses

JMP Genomics 5 (SAS Institute, Cary, NC, U.S.A.) software was used to create genetic distance matrices, perform principal component analysis (PCA), and graph global accession protein content. Variance components were calculated with the PROC GLIMMIX procedure using SAS software version 9.2 (SAS Institute). The model effects included the rice accessions, test locations, accession × location interactions, and replications, all considered as random factors. Of the 202 accessions used in the protein study, only the 182 *O. sativa* accessions that were successfully harvested at both the TX and AR locations were used for the population structure and association mapping analyses. The genetic distance matrix used for PCA and subsequent association analyses was based on Gower's similarity (Gower 1966), with PCA performed on the similarity correlation matrix.

#### Association Mapping

For the association analysis, rare marker alleles, occurring at a frequency less than 3.3% among all the accessions for any one marker, were converted into missing data, which included all heterozygous genotypes. Subsequently, eight simple sequence repeat markers were identified that had 33% or more accessions with missing allele calls, and these markers were excluded from the association analyses. A total of 157 markers distributed approximately every 10 cM across the rice genome were used for association mapping. TASSEL 3.0 (Bradbury et al 2007) software was used for the association analyses with the mixed linear model option (Yu et al 2006; Zhang et al 2010). Values from the first three PCA dimensions were used as population structure (Q) input for association mapping. The similarity matrix used in PCA was multiplied by two and used as a kinship matrix (K) in the association analysis (Buckler Lab at Cornell University 2011).

Protein concentration measurements were determined in duplicate from each of the two field replicates from the Stuttgart, AR,



**Fig. 2.** Protein content of mini-core accessions of rice and rice-related species depicted by latitude and longitude of accession geographic origin.

and Beaumont, TX, locations and were used to calculate the mean protein concentration for each location. Accession means at each location were analyzed for marker associations both independently and when averaged over locations.

## RESULTS AND DISCUSSION

### Molecular Substructure of Mini-Core Collection

The PCA of genetic marker variation in the 182 rice mini-core accessions grown at both locations showed the accessions separating into five genetic groups of *O. sativa*: indica ( $n = 56$ ), aus ( $n = 37$ ), temperate japonica ( $n = 32$ ), tropical japonica ( $n = 31$ ), and aromatic ( $n = 5$ ) plus their admixtures ( $n = 21$ ) (Fig. 1). The admixtures were similar to those reported by Agrama et al (2010), indicating that these accessions are a result of intermating between rice subpopulations. This ances-

tral relationship and subpopulation structure among the mini-core accessions confirmed the five main rice subpopulations previously seen in other genetic structure studies of rice (Garris et al 2005; Agrama et al 2010). Determining the genetic structure of accessions analyzed in association mapping studies helps ensure that false marker-trait associations are not accepted, which can be a common problem if population substructure is not considered (Yu et al 2006; Ersöz et al 2007; Zhang et al 2010). This information is also useful to breeders for designing crosses within subpopulations to change protein concentration while minimizing genetic variability in other genomic regions.

### Protein Concentration Analyses

Among the 202 mini-core accessions grown in either of the two locations, genetics (i.e., accessions) contributed the most to the

**TABLE III**  
Mean Protein Contents of 202 Mini-Core Accessions of Rice and Rice-Related Species, Averaged over Two Laboratory Replications  
Taken on Two Field Replications Grown in Two U.S. Locations (Arkansas and Texas)

ACNO <sup>a</sup>	Region	Ancestry <sup>b</sup>	Protein (%) <sup>c</sup>	SD <sup>d</sup>	ACNO <sup>a</sup>	Region	Ancestry <sup>b</sup>	Protein (%) <sup>c</sup>	SD <sup>d</sup>
PI 402673	Southeast Asia	Admix	6.1	0.56	PI 412790	Southeast Asia	Aus	7.1	0.52
PI 400780	Africa	Admix	6.2	0.80	PI 353723	Southeast Asia	Aus	7.2	0.97
PI 223612	South America	Admix	6.2	0.39	PI 402689	Southeast Asia	Aus	7.2	0.95
PI 434614	Southeast Asia	Admix	6.2	0.92	PI 431092	Mideast	Aus	7.2	0.39
PI 430936	Southeast Asia	Admix	6.3	0.27	PI 353722	Southeast Asia	Aus	7.2	0.99
PI 400771	Africa	Admix	6.3	0.89	PI 373403	Southeast Asia	Aus	7.2	0.87
PI 412811	Southeast Asia	Admix	6.9	0.92	PI 373340	Southeast Asia	Aus	7.2	0.41
PI 596818	Southeast Asia	Admix	6.9	0.43	PI 389863	Southeast Asia	Aus	7.2	1.11
PI 431128	Southeast Asia	Admix	6.9	0.20	PI 240638	Southeast Asia	Aus	7.2	0.64
PI 154435	China	Admix	7.3	0.51	PI 392677	Southeast Asia	Aus	7.4	0.59
PI 198134	Southeast Asia	Admix	7.6	0.34	PI 352687	North Pacific	Aus	7.4	1.27
PI 549224	Southeast Asia	Admix	7.7	0.77	PI 403469	Africa	Aus	7.5	1.22
CI 9049	Central America	Admix	7.8	1.02	PI 584620	Central Asia	Aus	7.5	1.34
CI 9723	Southeast Asia	Admix	7.9	0.10	PI 403214	Southeast Asia	Aus	7.7	1.31
PI 430979	Mideast	Admix	8.1	0.69	PI 392630	Africa	Aus	7.7	0.66
PI 215970	China	Admix	8.1	0.47	PI 549253	Southeast Asia	Aus	7.7	0.93
PI 245071	China	Admix	8.2	1.39	PI 403114	Southeast Asia	Aus	7.9	1.31
PI 584555	Southeast Asia	Admix	8.3	1.19	PI 393112	Southeast Asia	Aus	8.1	0.58
PI 373232	Southeast Asia	Admix	8.4	0.59	PI 430339	Oceania	Aus	8.2	0.91
PI 373249	Southeast Asia	Admix	8.4	1.53	PI 392768	Southeast Asia	Aus	9.6	1.43
PI 430909	Southeast Asia	Admix	8.5	1.43	PI 389923	China	IND	5.4	0.65
PI 400672	Western Europe	Admix	8.7	0.99	PI 389234	Southeast Asia	IND	5.7	1.03
PI 154481	South Pacific	Admix	9.3	0.89	PI 403422	Africa	IND	5.8	1.14
PI 256340	Southeast Asia	ARO	7.8	0.62	PI 160700	China	IND	5.9	0.20
PI 277414	Southeast Asia	ARO	8.0	1.35	PI 373335	Southeast Asia	IND	5.9	0.76
PI 406073	South America	ARO	8.4	0.97	PI 400398	China	IND	5.9	0.89
PI 584625	Central Asia	ARO	9.2	0.61	PI 67153	Southeast Asia	IND	6.3	0.59
PI 439669	Central Asia	ARO	9.3	0.62	PI 596941	Southeast Asia	IND	6.3	0.57
PI 431310	South Pacific	ARO	10.1	0.25	PI 493131	Southeast Asia	IND	6.3	0.72
PI 277415	Southeast Asia	Aus	5.6	0.36	PI 403289	Southeast Asia	IND	6.3	1.04
PI 392637	Africa	Aus	5.9	0.44	PI 596990	South Pacific	IND	6.4	0.57
PI 392694	Southeast Asia	Aus	5.9	1.07	PI 392605	Southeast Asia	IND	6.5	1.10
PI 402983	South Pacific	Aus	6.2	0.98	PI 431210	Central Asia	IND	6.6	0.26
PI 584566	Southeast Asia	Aus	6.2	0.87	PI 388304	North Pacific	IND	6.6	1.11
PI 388436	South America	Aus	6.5	0.66	PI 391218	South America	IND	6.7	1.10
PI 385697	Southeast Asia	Aus	6.5	0.64	PI 497682	South Pacific	IND	6.7	0.85
PI 584564	Mideast	Aus	6.6	1.22	CI 7155	South Pacific	IND	6.9	0.59
PI 439024	Southeast Asia	Aus	6.7	0.25	PI 615022	China	IND	6.9	1.12
PI 418207	Africa	Aus	6.7	0.23	CI 12492	China	IND	6.9	0.58
PI 127076	Southeast Asia	Aus	6.8	1.25	PI 160530	China	IND	6.9	0.62
PI 433799	Africa	Aus	6.9	0.28	PI 464597	South Pacific	IND	6.9	0.63
PI 385826	Southeast Asia	Aus	6.9	1.06	PI 431201	Central Asia	IND	6.9	0.73
PI 373341	Southeast Asia	Aus	6.9	1.00	PI 432578	South Pacific	IND	7.0	1.28
PI 373347	Southeast Asia	Aus	6.9	1.04	PI 389069	China	IND	7.0	0.92
PI 431204	Central Asia	Aus	7.1	1.23	PI 602654	Central America	IND	7.0	0.81
PI 403082	Southeast Asia	Aus	7.1	0.75	PI 417820	South Pacific	IND	7.1	1.14
PI 431172	South America	Aus	7.1	0.94	PI 391264	Southeast Asia	IND	7.1	1.03

(continued on facing page)

<sup>a</sup> ACNO = accession number. ACNO and region are as designated by Agrama et al (2010).

<sup>b</sup> ARO = aromatic; IND = indica; TEJ = temperate japonica; TRJ = tropical japonica; and admix = admixture (genetically interbred).

<sup>c</sup> Protein means were calculated using two technical replications by two field replications by two locations; average = 7.5%, minimum = 5.4%, and maximum = 11.9%.

<sup>d</sup> SD = standard deviation of two technical replications by two field replications by two locations.

variation in protein concentration, accounting for 40.1% of the total variation (Table II). Although the proportion of variation owing to location was 8.0% for AR and 7.0% for TX, the variation owing to location was not significant (Table II). However, the variation owing to accession × location interaction was significant, accounting for 22.3% of the protein concentration variation (Table II), suggesting the importance of conducting protein analysis studies across multiple environments.

The protein concentrations for all 202 accessions, globally depicted in Figure 2, averaged 7.5% and ranged from 5.4% for PI 389923, an indica from China, to 11.9% for PI 597033, a temperate japonica from Western Europe (Table III). The range of protein concentration in our study is smaller than that reported for the International Rice Research Institute world collection ( $n = 17,587$  cultivars), which had a range of 4.3–18.2% and a mean of 9.5% (Gomez 1979). Protein concentration values among the rice sub-

populations and rice-related species, along with geographic summaries, are shown in Table IV. The temperate japonica accessions of the mini-core had the broadest protein range (5.8–11.9%) among the subpopulations. For the rice-related species, the *O. glaberrima* accessions had a lower mean protein content (6.2%) compared with *O. sativa*, whereas the protein content of the one *O. rufipogon* accession (8.0%) was higher than the *O. sativa* mean. The protein concentrations seen in these species are similar to those found by Ignacio and Juliano (1968), who reported protein ranges of 8.9–11.7% for *O. glaberrima*, 8.3–13.8% for *O. rufipogon*, and 7.3–13.6% for *O. sativa*. Wide variation in protein concentration was observed in all genetic subpopulations of *O. sativa*. The means of the aromatic accessions (8.8%) and temperate japonica accessions (8.2%) were greater than the general mean of the mini-core, whereas the indica and tropical japonica mean values were similar (7.4% for both) to that of the mini-core

TABLE III (continued from facing page)

ACNO <sup>a</sup>	Region	Ancestry <sup>b</sup>	Protein (%) <sup>c</sup>	SD <sup>d</sup>	ACNO <sup>a</sup>	Region	Ancestry <sup>b</sup>	Protein (%) <sup>c</sup>	SD <sup>d</sup>
PI 373536	Southeast Asia	IND	7.1	0.35	PI 615198	China	TEJ	8.1	0.33
PI 420960	South America	IND	7.2	0.19	PI 266122	Eastern Europe	TEJ	8.2	0.93
PI 391214	Western Europe	IND	7.2	1.11	PI 268003	Western Europe	TEJ	8.2	1.02
PI 596902	Southeast Asia	IND	7.2	0.49	PI 439683	Eastern Europe	TEJ	8.6	0.66
PI 389267	Southeast Asia	IND	7.2	1.41	PI 584632	Eastern Europe	TEJ	8.7	0.84
PI 264242	Central America	IND	7.3	0.93	PI 215478	Western Europe	TEJ	8.8	0.21
PI 393180	Africa	IND	7.3	0.81	PI 584629	Central Asia	TEJ	8.8	0.19
PI 615033	China	IND	7.3	1.06	PI 291539	Western Europe	TEJ	9.0	0.68
PI 392813	South America	IND	7.3	1.06	PI 291430	Central Asia	TEJ	9.1	0.63
PI 399748	South Pacific	IND	7.3	0.94	PI 281630	North Pacific	TEJ	9.3	0.73
PI 608431	Southeast Asia	IND	7.4	0.95	PI 189460	Western Europe	TEJ	9.6	1.07
PI 614989	China	IND	7.4	0.90	PI 439674	Eastern Europe	TEJ	10.3	0.38
CI 7404	China	IND	7.4	0.83	PI 265110	Eastern Europe	TEJ	10.5	0.44
PI 393070	Southeast Asia	IND	7.5	0.50	PI 282173	Eastern Europe	TEJ	11.1	0.45
PI 549215	Southeast Asia	IND	7.6	0.10	PI 245694	North Pacific	TEJ	11.4	0.28
PI 430740	Southeast Asia	IND	7.7	2.20	PI 597033	Eastern Europe	TEJ	11.9	0.08
PI 393292	Southeast Asia	IND	7.7	2.37	PI 199553	Central America	TRJ	6.2	0.37
PI 373795	Southeast Asia	IND	7.8	0.21	PI 389876	South Pacific	TRJ	6.3	0.48
PI 389933	Southeast Asia	IND	7.9	1.21	CI 2169	Central America	TRJ	6.5	0.32
PI 431499	Southeast Asia	IND	7.9	0.45	PI 291608	South America	TRJ	6.7	0.45
PI 389960	Southeast Asia	IND	8.0	1.68	PI 37215	South America	TRJ	6.7	0.36
PI 389037	China	IND	8.0	0.55	PI 469300	South Pacific	TRJ	6.8	0.85
PI 281914	Southeast Asia	IND	8.4	2.44	CI 11030	North America	TRJ	6.8	0.55
PI 214077	Central America	IND	8.4	1.19	PI 155990	Oceania	TRJ	6.8	0.31
PI 400607	China	IND	8.5	2.35	CI 8913	Central America	TRJ	6.8	0.37
PI 263813	South America	IND	8.8	2.67	PI 220214	Central America	TRJ	6.9	0.83
PI 321183	South Pacific	IND	8.8	0.89	PI 434632	Africa	TRJ	6.9	0.25
PI 408406	South Pacific	IND	8.9	2.46	CI 2490	South Pacific	TRJ	7.0	0.30
PI 430254	Central America	IND	9.0	2.10	PI 161567	South America	TRJ	7.0	0.96
PI 389150	Southeast Asia	IND	9.0	1.09	PI 373781	Oceania	TRJ	7.1	0.70
PI 615192	China	IND	9.0	0.56	PI 282769	Africa	TRJ	7.1	0.66
PI 202864	Central America	IND	9.1	2.37	PI 54344	Southeast Asia	TRJ	7.2	0.66
PI 208452	Southeast Asia	IND	9.1	0.12	PI 282768	Africa	TRJ	7.2	0.47
PI 389360	China	IND	9.5	0.66	PI 561734	North America	TRJ	7.3	0.40
PI 226313	North America	IND	10.8	0.69	CI 9032	North America	TRJ	7.3	0.66
PI 154531	North Pacific	TEJ	5.8	0.26	PI 373899	Eastern Europe	TRJ	7.3	0.60
PI 177224	Mideast	TEJ	6.3	0.59	PI 373816	South Pacific	TRJ	7.5	1.02
PI 514663	North Pacific	TEJ	6.4	0.36	CI 9979	North America	TRJ	7.5	0.82
PI 388427	South America	TEJ	6.5	0.23	PI 154464	China	TRJ	7.6	0.20
PI 162113	North Pacific	TEJ	6.5	0.44	PI 373771	Oceania	TRJ	7.7	0.77
PI 373761	Oceania	TEJ	6.6	0.70	PI 602637	Africa	TRJ	7.7	0.33
PI 373939	Southeast Asia	TEJ	6.7	0.33	PI 430397	Africa	TRJ	7.9	2.05
PI 346827	South America	TEJ	6.8	0.58	PI 430387	Africa	TRJ	8.0	0.72
PI 182254	Mideast	TEJ	7.1	0.20	PI 373786	Oceania	TRJ	8.2	0.49
PI 494105	North America	TEJ	7.1	0.77	PI 373820	South Pacific	TRJ	8.4	0.68
PI 402789	Oceania	TEJ	7.1	1.97	PI 584567	Southeast Asia	TRJ	8.5	0.51
PI 168946	Western Europe	TEJ	7.3	1.02	CI 12153	South Pacific	TRJ	8.5	0.64
PI 184506	North Pacific	TEJ	7.5	0.69	PI 585042	South America	TRJ	8.7	1.14
PI 226204	North Pacific	TEJ	7.5	0.70	PI 373798	South Pacific	TRJ	9.1	0.50
PI 157385	North Pacific	TEJ	7.7	0.53	PI 269630	Central America	<i>Oryza glaberrima</i>	5.8	0.30
PI 602606	Eastern Europe	TEJ	7.8	0.69	PI 450365	Africa	<i>O. glaberrima</i>	6.3	1.04
PI 402794	Western Europe	TEJ	7.9	0.89	PI 450353	Africa	<i>O. glaberrima</i>	6.3	0.48
PI 267996	Western Europe	TEJ	7.9	0.42	PI 450396	Africa	<i>O. glaberrima</i>	6.5	0.89
PI 584624	Central Asia	TEJ	7.9	0.68	PI 346371	South America	<i>O. rufipogon</i>	8.0	0.28

mean. The accessions native to the Eastern European region ( $n = 9$ ) had the highest mean value concentration (9.4%), whereas the accessions native to Africa had the lowest. All the accessions with a protein concentration greater than 9.5%, except PI 226313, 245694, 392768, and 431310, were native to Eastern and Western Europe ( $n = 18$ ), and 15 of the accessions from those regions were temperate japonica and had values above the mini-core mean. In contrast, all of the rice accessions from Africa ( $n = 18$ ) averaged below 8.0%, and 14 were below the mini-core mean value. Although we only analyzed the protein content in brown rice, we

would expect to see similar protein content for milled rice samples, given the high correlation of these two measurements ( $r = 0.96\text{--}0.97$ ) seen in prior studies (Juliano et al 1964; Ellis et al 1986). Additionally, the protein content of brown rice is more standardized across a broad range of germplasm, particularly as the protein content of milled rice is notably affected by the degree of milling, which can be difficult to control (Resurreccion et al 1979; Pedersen and Eggum 1983).

#### Association Mapping

Results of the association mapping analyses showed significant protein concentration associations ( $P < 0.01$ ), with four markers in AR and five markers in TX (Table V). None of the significant markers were identical between the AR and TX locations, suggesting that environment influences protein concentration in the rice grain. Interestingly, all marker alleles showing the greatest amount of change in protein concentration for the marker-trait associations in AR had a positive effect on protein concentration, whereas all the major marker alleles identified in the TX location had a negative effect on protein concentration. The relatively modest significance levels among the marker-trait associations ( $-\log_{10}P$  values = 2.01–3.71) are fairly typical for those seen in quantitative trait locus (QTL) mapping studies of protein concentration (Yoshida et al 2002; Wada et al 2006; Kobayashi et al 2008; Wang et al 2008; Lu et al 2009). The lack of notably high marker significance levels may also partially result from the limited number of markers and germplasm analyzed in this study, particularly in comparison to recent studies in rice surveying roughly 44,000 single nucleotide polymorphism markers in 413 rice accessions (Famoso et al 2011; Zhao et al 2011), suggesting that more markers and accessions may be appropriate for subsequent studies on marker-protein association analyses.

In AR, two of the significant markers occurred at or near gene regions likely to be involved with protein concentration regulation. Two significant markers were novel and had not been previously reported. RM125 had the most significant  $P$  value, and its allele with the largest effect had a positive effect on protein concentration and was observed in 76 accessions. RM125 is on rice chromosome 7 at reference map position 5.5 Mb and occurs within 1 Mb of three known prolamin precursor genes, a low-molecular-weight globulin storage protein gene, and three albumin storage protein genes. RM302 on chromosome 1 at position 32.9 Mb occurs within 1 Mb of a predicted glutelin precursor gene, and this marker had a net positive effect on protein concen-

**TABLE IV**  
Statistical Summary of Brown Rice Protein Content (%)  
Among 202 Diverse Accessions of Rice and Rice-Related Species  
from the U.S. Department of Agriculture Mini-Core Germplasm  
Collection, Grouped According to Population Ancestry  
and Geographic Region of Origin

Accession Grouping	<i>n</i>	Mean (%)	Range (%)
Ancestry			
<i>Oryza sativa</i>	197	7.45	
Aromatic	6	8.80	7.76–10.13
Aus	38	7.15	5.61–9.57
Indica	62	7.39	5.44–10.83
Temperate japonica	35	8.18	5.76–11.94
Tropical japonica	34	7.38	6.21–9.15
Admixture	21	7.38	6.10–8.70
<i>O. glaberrima</i>	4	6.23	5.84–6.50
<i>O. rufipogon</i>	1	7.97	...
Region			
North Pacific	10	7.61	5.76–11.43
South Pacific	19	7.64	6.16–10.13
Oceania	7	7.39	6.59–8.25
China	19	7.43	5.44–9.51
Southeast Asia	19	7.68	5.74–9.57
Southern Asia	46	7.12	5.61–9.10
Central Asia	9	8.05	6.56–9.29
Mideast	5	7.03	6.29–8.05
Africa	15/18 <sup>a</sup>	7.01/6.90 <sup>a</sup>	5.83–7.97 <sup>b</sup>
Eastern Europe	9	9.38	7.34–11.94
Western Europe	9	8.29	7.18–9.64
South America	14/15 <sup>a</sup>	7.18/7.23 <sup>a</sup>	6.18–8.78 <sup>b</sup>
Central America	10/11 <sup>a</sup>	7.50/7.35 <sup>a</sup>	6.21–9.09 <sup>b</sup>
North America	6	7.83	6.79–10.83

<sup>a</sup> Values shown are those for the *O. sativa* accessions only/*O. sativa* plus related species accessions.

<sup>b</sup> Protein content range of *O. sativa* subspecies only (not related species).

**TABLE V**  
Mixed Linear Model Association Mapping Results for Arkansas, Texas, and the Combined Location Study<sup>a</sup>

Location	Locus	Chromosome	Position (Mb)	<i>P</i> Value	$-\log_{10}P$	Major Allele	Effect	Observed Genotypes
Arkansas	RM125 <sup>b</sup>	7	5.5	0.0009	3.02	126	2.08	76
	RM302 <sup>b</sup>	1	32.9	0.0049	2.31	117	1.10	16
	RM471	4	18.8	0.0074	2.13	113	1.78	8
	RM341	2	19.4	0.0098	2.01	171	1.42	31
Texas	RM15123	3	15.8	0.0002	3.71	380	-1.45	12
	RM24011 <sup>c</sup>	9	9.4	0.0006	3.20	420	-1.42	13
	RM541 <sup>b</sup>	6	19.5	0.0014	2.85	161	-1.14	9
	RM427 <sup>b</sup>	7	2.7	0.0037	2.43	186	-0.70	72
	RM413	5	2.2	0.0050	2.30	77	-0.83	63
Combined	RM15123	3	15.8	0.0006	3.20	380	-1.25	12
	RM541 <sup>b</sup>	6	19.5	0.0013	2.88	165	1.25	9
	RM24011 <sup>c</sup>	9	9.4	0.0017	2.78	420	-1.28	13
	RM471	4	18.8	0.0034	2.46	113	1.78	8
	RM125 <sup>b</sup>	7	5.5	0.0051	2.29	126	1.31	76
	RM3558	4	22.8	0.0066	2.18	104	-1.40	7

<sup>a</sup> Marker-trait associations with significant  $P$  values less than 0.01 are shown, along with chromosomal location and position (in Mb). Marker effects are shown for the major allele (alleles having the greatest effect on protein content) for each locus. Effects are shown as either a negative value indicating that the presence of the allele was associated with decreased protein content or a positive value indicating that the presence of the allele was associated with increased protein content. The number of observed genotypes is the numbers of times the allele occurred in a homozygous state in the 182 lines tested.

<sup>b</sup> Marker occurs within 1.5 Mb of protein candidate gene.

<sup>c</sup> Occurs within an identified quantitative trait locus region for increased protein content.

tration. RM471 (chromosome 4 at 18.8 Mb) and RM341 (chromosome 2 at 19.4 Mb) were novel marker-trait associations identified in this study, and all of the alleles had either no effect or a net positive effect on protein concentration.

Of the five significant marker-trait associations identified in the TX study, three occur within 1.5 Mb of previously identified protein content related QTLs or predicted protein concentration regulatory genes. RM541 (chromosome 6 at 19.5 Mb) occurs within 1.5 Mb of a predicted prolamin precursor gene and a previously identified QTL (Lu et al 2009). RM427 (chromosome 7 at 2.7 Mb) occurs within 1.5 Mb of a previously identified QTL associated with reduced protein content (Shi et al 2009). RM24011 (chromosome 9 at 9.4 Mb) occurs within an identified QTL region for increased protein content (Wada et al 2006). RM15123 (chromosome 3 at 15.8 Mb), and RM413 (chromosome 5 at 2.2 Mb) are novel marker-trait associations identified in this study, and all of their major alleles had negative effects on protein concentration. RM15123 had the most significant *P* value as well as the largest negative effect on protein concentration of all the marker-trait associations identified.

Although the accession × location term was statistically significant (Table II), the location main effect was not. Thus, a combined location association analysis was performed by using an average of the means of both AR and TX (Table V). Two significant markers identified in the AR analysis, three in the TX analysis, and one new marker were identified as significant (*P* < 0.01) in the combined location analysis. Markers RM15123, RM541, and RM24011 had the most significant *P* values in both the combined analysis and the TX study. In the combined location analysis, RM541, which is near a previously identified prolamin QTL and predicted prolamin precursor (Lu et al 2009), had a large positive allele effect on protein, whereas in the TX study a different major allele was identified for this marker, which had an overall negative effect. This difference suggests that the effect this marker has on protein concentration is dependent on the environment. The same alleles for markers RM15123 and RM24011 were observed to have a consistent negative effect in both the combined analysis and at TX. RM471 and RM125 were both present in the AR study, and combined location analysis and the major alleles consistently had a positive effect. RM125, RM302, and RM427 are within 1.5 Mb of protein candidate genes (Rice Genome Annotation Project, rice.plantbiology.msu.edu). RM3558, which was not detected in either the AR or TX study, was identified in the combined location analysis with the predominant allele having a negative impact on protein. RM471 (chromosome 4 at 18.8 Mb) and RM3558 (chromosome 4 at 22.8 Mb) are close to each other, suggesting that this region contains a novel protein concentration QTL.

## CONCLUSIONS

The protein concentration of rice was variable across all subpopulations of rice, which allows breeders to select accessions with high or low protein content for crossing within the same subpopulation and thus alleviating issues associated with wide crosses, such as sterility and incompatibility. In addition, accessions from within various global regions differed widely for protein concentration, indicating that breeders have many options in using germplasm adapted to different growing environments. The accession × location effect played a large role in determining protein concentration, highlighting the importance of assessing gene × environment interactions in protein studies. Selecting accessions that are genetically the best suited for the location in which the crop will be grown will be an important factor to consider for breeding programs. Ten significant markers were identified in the association mapping studies; five were in the TX study, four were in the AR study, and one additional marker not identified in either the TX or AR location was identified in the com-

bined location analysis. Of these significant markers, five were located at or near genes known or predicted to be responsible for protein content or near regions previously identified as QTL involved in protein content of the rice grain. The remaining five are novel markers identified in this study. Two of these novel markers, RM471 and RM3558, occur in fairly close proximity to each other on chromosome 4, which suggests this region would be a good candidate region for further study. Association mapping of the USDA rice mini-core collection was an effective method for identifying new genetic markers and validating previously reported marker regions associated with protein content in rice.

## ACKNOWLEDGMENTS

The authors thank Melissa Jia, Tiffany Sookaserm, Heather Box, Yao Zhou, LaDuska Simpson, Curtis Kerns, Sarah Hendrix, Bill Luebke, Jodie Cammack, Kip Landry, Carl Henry, Jason Bonnette, and Piper Roberts for technical assistance.

## LITERATURE CITED

- AACC International. Approved Methods of Analysis, 11th Ed. Method 46-30.01. Crude protein—Combustion method. Approved November 8, 1995. AACC International: St. Paul, MN. <http://dx.doi.org/10.1094/AACCIntMethod-46-30.01>
- Agrama, H. A., Eizenga, G. C., and Yan, W. 2007. Association mapping of yield and its components in rice cultivars. *Mol. Breed.* 19:341–356.
- Agrama, H. A., Yan, W., Lee, F., Fjellstrom, R., Chen, M.-H., Jia, M., and McClung, A. M. 2009. Genetic assessment of a mini-core subset developed from the USDA rice genebank. *Crop Sci.* 49:1336–1346.
- Agrama, H. A., Yan, W., Jia, M., Fjellstrom, R., and McClung, A. M. 2010. Genetic structure associated with diversity and geographic distribution in the USDA rice world collection. *Nat. Sci.* 2:247–291.
- Bao, J. S., Corke, H., and Sun, M. 2006. Nucleotide diversity in starch synthase IIa and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 113:1171–1183.
- Bockelman, H. E., Dilday, R. H., Yan, W. G., and Wesenberg, D. M. 2003. Germplasm collection, preservation and utilization. Pages 597–625 in: *Rice Origin, History, Technology and Production*. C. W. Smith and R. H. Dilday, eds. Wiley: Hoboken, NJ.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635.
- Bresghello, F., and Sorrells, M. E. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genet.* 172:1165–1177.
- Bryant, R., Proctor, A., Hawkrige, M., Jackson, A., Yeater, K., Counce, P., Yan, W., McClung, A., and Fjellstrom, R. 2011. Genetic variation and association mapping of silica concentration in rice hulls using a germplasm collection. *Genetica* 139:1383–1398.
- Bryant, R., Anders, M., and McClung, A. 2012. Impact of production practices on physicochemical properties of rice grain quality. *J. Sci. Food Agric.* 92:564–569.
- Buckler Lab at Cornell University. 2011. User manual for TASSEL: Trait analysis by association, evolution and linkage. Ithaca, NY. Available at: [www.maizegenetics.net/tassel](http://www.maizegenetics.net/tassel)
- Champagne, E. T., Bett-Garber, K. L., Grimm, C. C., and McClung, A. M. 2007. Effects of organic fertility management on physicochemical properties and sensory quality of diverse rice cultivars. *Cereal Chem.* 84:320–327.
- Champagne, E. T., Bett-Garber, K. L., Thomson, J. L., and Fitzgerald, M. A. 2009. Unraveling the impact of nitrogen nutrition on cooked rice flavor and texture. *Cereal Chem.* 86:274–280.
- Chen, M.-H., Fjellstrom, R. G., Christensen, E. F., and Bergman, C. J. 2010. Development of three allele-specific codominant rice *Waxy* gene PCR markers suitable for marker-assisted selection of amylose content and paste viscosity. *Molec. Breed.* 26:513–523.
- Eggum, B. O. 1969. Evaluation of protein quality and the development of screening techniques. Pages 125–135 in: *New Approaches to Breeding for Improved Plant Protein*. International Atomic Energy Agency: Vienna.
- Eggum, B. O. 1977. Nutrition aspects of cereal proteins. Pages 349–369

- in: Genetic Diversity in Plants. A. Muhammed, R. Aksel, and R. C. von Borstel, eds. Plenum Press: New York, NY.
- Eggum, B. O. 1979. The nutritional value of rice in comparison with other cereals. Pages 91-111 in: Proceedings of the Workshop on Chemical Aspects of Rice Grain Quality. International Rice Research Institute: Los Baños, Laguna, Philippines.
- Ellis, J. R., Villareal, C. R., and Juliano, B. O. 1986. Protein content, distribution and retention during milling of brown rice. Qual. Plant.—Plant Foods Hum. Nutr. 36:17-26.
- Ersoz, E. S., Yu, J., and Buckler, E. 2007. Applications of linkage disequilibrium and association mapping in crop plants. Pages 97-119 in: Genomics-Assisted Crop Improvement. R. Varshney and R. Tuberosa, eds. Springer: Dordrecht, The Netherlands.
- Famoso, A. N., Zhao, K., Clark, R. T., Tung, C.-W., Wright, M. W., Bustamante, C., Kochian, L. V., and McCouch, S. R. 2011. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. PLoS Genet. 7:e1002221.
- Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., and McCouch, S. 2005. Genetic structure and diversity in *Oryza sativa* L. Genetics 169:1631-1638.
- Gomez, K. A. 1979. Effect of environment on protein and amylose content of rice. Pages 59-68 in: Proceedings of the Workshop on Chemical Aspects of Rice Grain Quality. International Rice Research Institute: Los Baños, Laguna, Philippines.
- Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53:325-338.
- He, Y., Ning, T., Xie, T., Qiu, Q., Zhang, L., Sun, Y., Jiang, D., Fu, K., Yin, F., Zhang, W., Shen, L., Wang, H., Li, J., Lin, Q., Sun, Y., Li, H., Zhu, Y., and Yang, D. 2011. Large-scale production of functional human serum albumin from transgenic rice seeds. Proc. Natl. Acad. Sci. USA 108:19078-19083.
- Hulbert, S. H., and Bennetzen, J. L. 1991. Recombination at the *Rpl* locus of maize. Molec. Gen. Genet. 226:377-382.
- Ignacio, C. I., and Juliano, B. O. 1968. Physicochemical properties of brown rice from *Oryza* species and hybrids. J. Agric. Food Chem. 16:125-127.
- Juliano, B. O., Cagampang, G. B., Cruz, L. J., and Santiago, R. G. 1964. Some physicochemical properties of rice in Southeast Asia. Cereal Chem. 41:275-286.
- Kawakatsu, T., Yamamoto, M. P., Hirose, S., Yano, M., and Takaiwa, F. 2008. Characterization of a new rice glutelin gene *GluD-1* expressed in the starchy endosperm. J. Exp. Bot. 59:4233-4245.
- Kawakatsu, T., Hirose, S., Yasuda, H., and Takaiwa, F. 2010. Reducing rice seed storage protein accumulation leads to changes in nutrient quality and storage organelle formation. Plant Physiol. 154:1842-1854.
- Kobayashi, S., Fukuta, Y., Takeda, H., Sato, T., and Osaki, M. 2008. Identification and characterization of genomic regions associated with nitrogen dynamics in rice plants (*Oryza sativa* L.). Breed. Sci. 58:113-120.
- Kraakman, A. T. W., Niks, R. E., Van den Berg, P., Stam, P., and Van Eeuwijk, F. A. 2004. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435-446.
- Li, X., Yan, W., Agrama, H., Hu, B., Jia, L., Jia, M., Jackson, A., Moldenhauer, K., McClung, A., and Wu, D. 2010. Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. Genetica 138:1221-1239.
- Little, R. R., and Dawson, E. H. 1960. Histology and histochemistry of raw and cooked rice kernels. Food Res. 25:611-622.
- Lu, K., Li, L., Zheng, X., Zhang, Z., Moua, T., and Hu, Z. 2009. Genetic dissection of amino acid content in rice grain. J. Sci. Food Agric. 89:2377-2382.
- Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hincliffe, E., Adams, J. L., Silverstone, A. L., and Drake, R. 2005. Improving the nutritional value of Golden Rice through increased pro-vitamin A content. Nat. Biotechnol. 23:482-487.
- Pedersen, B., and Eggum, B. O. 1983. The influence of milling on the nutritive value of flour from cereal grains. Qual. Plant.—Plant Foods Hum. Nutr. 33:267-278.
- Qu, L. Q., and Takaiwa, F. 2004. Evaluation of tissue specificity and expression strength of rice seed component gene promoters in transgenic rice. Plant Biotechnol. J. 2:113-125.
- Ravel, C., Praud, S., Murigneux, A., Linossier, L., Dardevet, M., Balfourier, F., Dufour, P., Brunel, D., and Charmet, G. 2006. Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study. Theor. Appl. Genet. 112:738-743.
- Resurreccion, A. P., Juliano, B. O., and Tanaka, Y. 1979. Nutrient content and distribution in milling fractions of rice grain. J. Sci. Food Agric. 30:475-481.
- Shi, C. H., Shi, Y., Lou, X. Y., Xu, H. M., Zheng, X., and Wu, J. G. 2009. Identification of endosperm and maternal plant QTLs for protein and lysine contents of rice across different environments. Crop Pasture Sci. 60:295-301.
- Shih, F. F. 2004. Rice proteins. Pages 143-162 in: Rice: Chemistry and Technology, 3rd Ed. E. T. Champagne, ed. American Association of Cereal Chemists: St. Paul, MN.
- Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and McCouch, S. 2006. Caught red-handed: *Rc* encodes a basic helix-loop-helix protein conditioning red pericarp in rice. Plant Cell 18:283-294.
- Tada, Y., Utsumi, S., and Takaiwa, F. 2003. Foreign gene products can be enhanced by introduction into low storage protein mutants. Plant Biotechnol. J. 1:411-422.
- Wada, T., Uchimura, Y., Ogata, T., Tsubone, M., and Matsue, Y. 2006. Mapping of QTLs for physicochemical properties in *japonica* rice. Breed Sci. 56:253-260.
- Wang, L., Zhong, M., Li, X., Yuan, D., Xu, Y., Liu, H., He, Y., Luo, L., and Zhang, Q. 2008. The QTL controlling amino acid content in grains of rice (*Oryza sativa*) are co-localized with regions involved in the amino acid metabolism pathway. Mol. Breed. 21:127-137.
- Wang, X., Fjellstrom, R., Jia, Y., Yan, W. G., Jia, M. H., Scheffler, B. E., Wu, D., Shu, Q., and McClung, A. 2010. Characterization of *Pi-ta* blast resistance gene in an international rice core collection. Plant Breed. 129:491-501.
- Wolff, I. A., ed. 1982. CRC Handbook of Processing and Utilization in Agriculture. Vol. II, Part I. Plant Products. CRC Press: Boca Raton, FL.
- Xin, Z., Velten, J. P., Oliver, M. J., and Burke, J. J. 2003. High-throughput DNA extraction method suitable for PCR. BioTechniques 34:802-826.
- Yan, W., Rutger, J. N., Bryant, R. J., Bockelman, H. E., Fjellstrom, R. G., Chen, M.-H., Tai, T. H., and McClung, A. M. 2007. Development and evaluation of a core subset of the USDA rice germplasm collection. Crop Sci. 47:869-878.
- Yoshida, S., Ikegami, M., Kuze, J., Sawada, K., Hashimoto, Z., Ishii, T., Nakamura, C., and Kamijima, O. 2002. QTL analysis for plant and grain characters of sake-brewing rice using a doubled haploid population. Breed. Sci. 52:309-317.
- Yu, J., Pressoir, G., Briggs, W. H., Vroh, B. L., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., and Buckler, E. S. 2006. A unified mixed model for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203-208.
- Zhang, Y. M., Mao, Y., Xie, C., Smith, H., Luo, L., and Xu, S. 2005. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (*Zea mays* L.). Genetics 169:2267-2275.
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., and Buckler, E. S. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355-360.
- Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., Islam, M. R., Reynolds, A., Mezey, J., McClung, A. M., Bustamante, C. D., and McCouch, S. R. 2011. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat. Commun. 2:467.

[Received September 28, 2012. Accepted May 2, 2013.]